

<https://helda.helsinki.fi>

Systematic survey of randomized trials evaluating the impact of alternative diagnostic strategies on patient-important outcomes

El Dib, Regina

2017-04

El Dib , R , Tikkinen , K A O , Akl , E A , Gomaa , H A , Mustafa , R A , Agarwal , A , Carpenter , C R , Zhang , Y , Jorge , E C , Almeida , R A M B , do Nascimento Junior , P , Doles , J V P , Mustafa , A A , Sadeghirad , B , Lopes , L C , Bergamaschi , C C , Suzumura , E A , Cardoso , M M A , Corrente , J E , Stone , S B , Schunemann , H J & Guyatt , G H 2017 , ' Systematic survey of randomized trials evaluating the impact of alternative diagnostic strategies on patient-important outcomes ' , Journal of Clinical Epidemiology , vol. 84 , pp. 61-69 . <https://doi.org/10.1016/j.jclinepi.2016.12.009>

<http://hdl.handle.net/10138/237024>

<https://doi.org/10.1016/j.jclinepi.2016.12.009>

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Systematic survey of randomized trials evaluating the impact of alternative diagnostic strategies on patient-important outcomes

Regina El Dib^{a,b,c,*}, Kari A.O. Tikkinen^d, Elie A. Akl^{e,f}, Huda A. Gomaa^g, Reem A. Mustafa^{h,i},
Arnav Agarwal^j, Christopher R. Carpenter^k, Yuchen Zhang^l, Eliane C. Jorge^m,
Ricardo A.M.B. Almeidaⁿ, Paulo do Nascimento Junior^a, Joao Vitor P. Doles^a,
Ahmad A. Mustafa^o, Behnam Sadeghirad^p, Luciane C. Lopes^q, Cristiane C. Bergamaschi^q,
Erica A. Suzumura^r, Marília M.A. Cardoso^a, José Eduardo Corrente^s,
Samuel B. Stone^t, Holger J. Schunemann^{h,u}, Gordon H. Guyatt^{h,v}

^aDepartment of Anaesthesiology, Botucatu Medical School, Unesp—Univ Estadual Paulista, Distrito de Rubião Júnior, s/n Univ Estadual Paulista, São Paulo 18618-970, Brazil

^bDepartment of Biosciences and Oral Diagnosis, Institute of Science and Technology, Unesp—Univ Estadual Paulista, Av. Eng. Francisco José Longo, 777 - Jardim Sao Dimas, São José dos Campos 12245-000, Brazil

^cMcMaster Institute of Urology, McMaster University, St. Josephs Healthcare, 50 Charlton Ave E, Room G344, Hamilton, Ontario L8N 4A6, Canada

^dDepartment of Urology, University of Helsinki and Helsinki University Hospital, Haartmaninkatu 4, Helsinki 00029, Finland

^eDepartment of Internal Medicine, American University of Beirut Medical Center, P.O. Box 11-0236, Riad-El-Solh, Beirut 1107 2020, Lebanon

^fDepartment of Health Research Methods, Evidence, and Impact (HEI), 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada

^gDepartment of Pharmacy, Tanta Chest Hospital, El Bahr St., Tanta, Gharbia 31527, Egypt

^hClinical Epidemiology and Biostatistics, Faculty of Health Sciences, HSC-3V43D 1200 Main Street West, Hamilton, Ontario L8N 3Z5, Canada

ⁱDepartment of Medicine, University of Missouri-Kansas City School of Medicine, M4-303, 2411 Holmes St., Kansas City, MO 64108-2792, USA

^jSchool of Medicine, University of Toronto, 1 King's College Circle, Toronto, Ontario M5S 1A8, Canada

^kDepartment of Emergency Medicine, Washington University School of Medicine in St. Louis, Campus Box 8072, 660 S. Euclid Avenue, St. Louis, MO 63110, USA

^lDepartment of Health Sciences, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada

^mDepartment of Ophthalmology, Otorhinolaryngology and Head and Neck Surgery, Botucatu Medical School, Unesp—Univ Estadual Paulista, Distrito de Rubião Júnior, s/n Univ Estadual Paulista, São Paulo 18618-970, Brazil

ⁿDepartment of Tropical Diseases and Imaging Diagnosis, Botucatu Medical School, Unesp—Univ Estadual Paulista, Distrito de Rubião Júnior, s/n Univ Estadual Paulista, São Paulo 18618-970, Brazil

^oSchool of Medicine, Jordan University of Science and Technology, P.O. Box 3030, Irbid 22110, Jordan

^pResearch Center for Modeling in Health, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

^qPharmaceutical Science Graduate Course, University of Sorocaba, Rodovia Raposo Tavares, Km 92, 5, Sorocaba, São Paulo 18023-000, Brazil

^rResearch Institute, Hospital do Coração (HCor), Abílio Soares Street 250, 12th Floor, São Paulo 04005-000, Brazil

^sBiostatistics Department, Biosciences Institute, Unesp—Univ Estadual Paulista, Distrito de Rubião Júnior, s/n Univ Estadual Paulista, São Paulo 18618-970, Brazil

^tFamily Medicine and CCFP [EM] Programmes, Northern Ontario School of Medicine, 955 Oliver Road, Thunder Bay, Ontario P7B 5E1, Canada

^uEmergency Department, Health Sciences North, 41 Ramsey Lake Rd, Sudbury, Ontario P3E 5J1, Canada

^vDepartment of Medicine, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4L8, Canada

Accepted 15 December 2016; Published online 4 January 2017

Abstract

Objectives: To provide a perspective on the current practice of randomized clinical trials (RCTs) of diagnostic strategies focusing on patient-important outcomes.

Study Design and Setting: We conducted a comprehensive search of MEDLINE and included RCTs published in full-text reports that evaluated alternative diagnostic strategies.

Funding: R.E.D. certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (e.g., employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending),

are the following: R.E.D. received a Brazilian Research Council (CNPq) scholarship (CNPq 310953/2015-4).

* Corresponding author. Tel.: +55-11-9-9999-6647; fax: +55-14-3880-1414.

E-mail address: eldib@ict.unesp.br (R. El Dib).

Results: Of 56,912 unique citations, we sampled 7,500 and included 103 eligible RCTs, therefore suggesting that MEDLINE includes approximately 781 diagnostic RCTs. The 103 eligible trials reported on: mortality ($n = 41$; 39.8%); morbidities ($n = 63$; 61.2%); symptoms/quality of life/functional status ($n = 14$; 13.6%); and on composite end points ($n = 10$; 9.7%). Of the studies that reported statistically significant results ($n = 12$; 11.6%), we judged 7 (58.3%) as at low risk of bias with respect to missing outcome data and 4 (33.3%) as at low risk of bias regarding blinding. Of the 41 RCTs that reported on mortality, only one (2.4%) reported statistically significant results. Of 63 RCTs addressing morbidity outcomes, 11 (17.5%) reported statistically significant results, all of which reported relative effects of greater than 20%.

Conclusion: RCTs of diagnostic tests are not uncommon, and sometimes suggest benefits on patient-important outcomes but often suffer from limitations in sample size and conduct. © 2017 Elsevier Inc. All rights reserved.

Keywords: Clinical trials; Diagnostic techniques and procedures; Accuracy; Alternative diagnostic strategies; Patient outcome; Evidence-based medicine

1. Introduction

Laboratory tests and medical imaging [1,2] are necessary for accurate diagnosis and constitute an essential component of patient management [3–5]. Clinicians often adopt tests for routine clinical use on the basis of diagnostic accuracy alone, implicitly assuming that use of accurate tests will improve outcomes: patients will live longer or live better. Even when tests are accurate, however, this may not be the case. A test may not provide incremental diagnostic information over and above inferences based on prior available information; even if a test provides incremental information, results may not change patient management or management may change, but the change may not improve outcome.

Thus, one can conceptualize a hierarchy of diagnostic evidence from that which addresses the capability to capture an image or quantify a laboratory finding; addresses diagnostic accuracy; evaluates test impact on patient management; and informs effects on patient-important outcomes [6–8]. This hierarchy implies that smaller subsets of patients will benefit from a test as researchers advance from simply measuring diagnostic accuracy to evaluating improvements in outcomes (Appendix Fig. 1 on the journal's Web site at www.elsevier.com) [9].

When, despite demonstration of test accuracy, patient benefit remains in doubt, randomized clinical trials (RCTs) that address the impact of alternative diagnostic strategies on patient-important outcomes are required [6–8,10–12]. This principle is well established for screening tests, and investigators have conducted many trials of screening tests. RCTs of test-and-treatment strategies are not, however, routinely performed, recognition of the importance of RCTs of diagnostic tests remains limited [10], and the RCTs thus far conducted remain poorly characterized.

We therefore conducted a systematic survey of diagnostic strategy RCTs to characterize their topic areas, population, setting, intervention and control groups, patient-important outcomes, risk of bias, and results.

2. Methods

2.1. Eligibility criteria

We included studies that met the following criteria:

- i) Randomized control trial.
- ii) Published in full-text report with no language restrictions.
- iii) Assessed alternative diagnostic tests or strategies (for instance, test A vs. test B or test A vs. no test). We defined “diagnostic studies” as those that evaluate tests used for diagnosis in patients presenting to any medical setting with symptoms or problems suggesting they may have a target condition. Test results in such situations either aim to decrease or increase the probability the target condition exists.
- iv) Examined the impact of the diagnostic strategies being evaluated on at least one patient-important outcome.

We excluded studies meeting the following criteria:

- v) Crossover studies.
- vi) Studies in which the only patient-important outcome measured was cost.
- vii) Studies in which the only patient-important outcomes measured were adverse effects of the testing procedure.
- viii) Screening studies (i.e., evaluating tests undertaken when patients have no symptoms or problems suggesting they may have a target condition).
- ix) Monitoring studies (patients already have the diagnosis of the condition of interest, and tests are being used to assess degree of improvement or deterioration).
- x) Studies focused exclusively on diagnostic test accuracy that did not report impact on patient-important outcomes.

2.2. Search strategy

An experienced research librarian searched in MEDLINE via OVID (1946 to December 1, 2013) using a comprehensive search strategy including both subject

What is new?**Key findings**

- Diagnostic tests are crucial for optimal patient management.
- Diagnostic laboratory or imaging strategies, sometimes invasive and often expensive, are often adopted based only on their diagnostic accuracy.
- Although accuracy alone can sometimes allow inferences regarding patient benefit, even accurate tests may not improve patient-important outcomes and therefore require randomized clinical trials (RCTs) of alternative diagnostic strategies to establish benefit.

What this adds to what was known?

- We provided a perspective on the current practice of RCTs of diagnostic strategies focusing on patient-important outcomes.

What is the implication and what should change now?

- Our results demonstrate the feasibility of RCTs of diagnostic tests, and their potential, when adequately powered and rigorously planned and carried out, of demonstrating benefit of one strategy over another on patient-important outcomes.
- There is a need for better reporting of diagnostic RCTs.

headings and free text words and an exhaustive list of synonyms for clinical trials and diagnosis. Known potentially eligible studies previously identified served as a guide to our search [13–17]. Appendix Table 1 on the journal's Web site at www.elsevier.com presents the detailed search strategy.

The search yielded 56,912 unique potentially eligible citations that we randomly sampled in batches of 2,500 by a computer-generated random number table. We obtained and evaluated full text of citations that appeared they might meet eligibility criteria. We sampled repeatedly until we achieved our target sample size of 100; three such samples were required.

2.3. Selection process and data extraction

We developed standardized, pilot-tested forms together with detailed instructions for screening of abstracts and full text, risk of bias assessments, and data collection. Reviewers conducted pilot screening and data extraction exercises. Pairs of reviewers, independently and in duplicate, screened study reports for eligibility,

assessed risk of bias, and abstracted data. Reviewers resolved disagreements through discussion and resolved remaining disagreements through consultation with one of four adjudicators (R.E.D., K.A.O.T., E.A.A., and G.H.G.). We abstracted the following characteristics of the eligible RCTs:

- Journal name; medical content area, single or multi-centre, and country.
- Characteristics of the population including the condition/disease under study and number of patients per arm.
- Characteristics of the interventions and control groups and duration of follow-up. We classified tests into the following major categories:
 - Diagnostic imaging (e.g., conventional x-ray, computed tomography, magnetic resonance imaging, ultrasound, and echocardiography).
 - Laboratory tests (e.g., blood, urine, or body tissues including biopsy).
 - Genetic tests (e.g., prenatal testing and DNA testing).
 - Molecular tests (e.g., nucleic acid amplification techniques).
 - Measurements of function or performance such as physical examination (e.g., blood pressure) and treadmill test.
 - Symptoms/drug challenging (e.g., injection of flumazenil for coma with suspected poisoning, allergy tests).
 - Other tests (e.g., polysomnography, laparoscopy).

We classified patient-important outcomes according to a previously established hierarchy of outcomes as mortality, morbidity, and symptoms/quality of life (QoL)/functional status [18] (Appendix Table 2 on the journal's Web site at www.elsevier.com). We recorded the point estimates [number of events and proportions, measure of relative effects (risk relative {RR}, and hazard ratio {HR}), *P* value, and confidence intervals (CIs)] for the most important outcome from each category (mortality, morbidity, and symptoms/QoL/functional status) as judged by the pair of reviewers. When the included studies did not report relative effects but provided sufficient information, we calculated relative risks and associated CIs. We reported composite end points only if authors explicitly designated them as primary outcomes. We also determined if the study included a strict protocol linking test results to a particular course of action, documented how many included sample size calculations, and of those that did, how many were powered to detect a relative risk reduction of 35% or less.

2.4. Risk of bias

Reviewers assessed risk of bias using a modified version of the Cochrane Collaboration's tool (<http://distillercer>).

com/resources/) [19]. We evaluated each trial according to six criteria: concealment, blinding, loss to follow-up, analysis of patients as randomized, selective outcome reporting, and other sources of bias. For incomplete outcome data in individual studies, we stipulated as low risk of bias loss to follow-up of less than 10% and a difference of less than 5% in missing data between the studied groups.

2.5. Statistical analysis

We estimated the number of diagnostic RCTs by multiplying the number of eligible studies in our sample by the ratio of the total number of citations identified (56,912) to the number sampled (7,500). We considered studies in which *P* values were less than 0.05 as statistically significant. We collected data on whether the studies that did not report statistically significant differences were adequately powered. Our power calculation specified a 25% relative risk reduction with an alpha of 0.05, beta of 0.2, event rate from each studies' control group, and a two-tailed hypothesis test.

2.6. Sample size

We desired a CI around all proportions of $\pm 10\%$; the widest CI occurs when the proportion is 0.5, and 100 studies would be required to meet this requirement.

3. Results

3.1. Study selection

Of 60,211 references identified, 56,912 titles and abstracts remained after removal of duplicates; from these, we randomly sampled 7,500. From these 7,500, we identified 176 potentially eligible RCTs of which 103 proved eligible after full-text review (Fig. 1). Given that, of 7,500 citations, 103 proved eligible, we estimate that MEDLINE to the end of 2013 includes approximately 781 diagnostic RCTs, with a 95% CI from 672 to 928.

3.2. Study characteristics

Few diagnostic RCTs were published prior to 2000 (Fig. 2); the number in each subsequent year is similar (Fig. 2). The articles were published in a wide variety of journals [44 journals; of the 103 RCTs, no journal published more than 8, (Appendix Table 3 on the journal's Web site at www.elsevier.com)] and addressed a wide variety of target diseases or conditions (Appendix Table 4 on the journal's Web site at www.elsevier.com) from many medical areas—most commonly cardiology ($n = 27$; 26.2%) and surgery ($n = 21$; 20.4%) (Fig. 3). Of the 103 studies, 53 (51.4%) were single center studies and almost all were two-arm studies ($n = 98$; 95.1%).

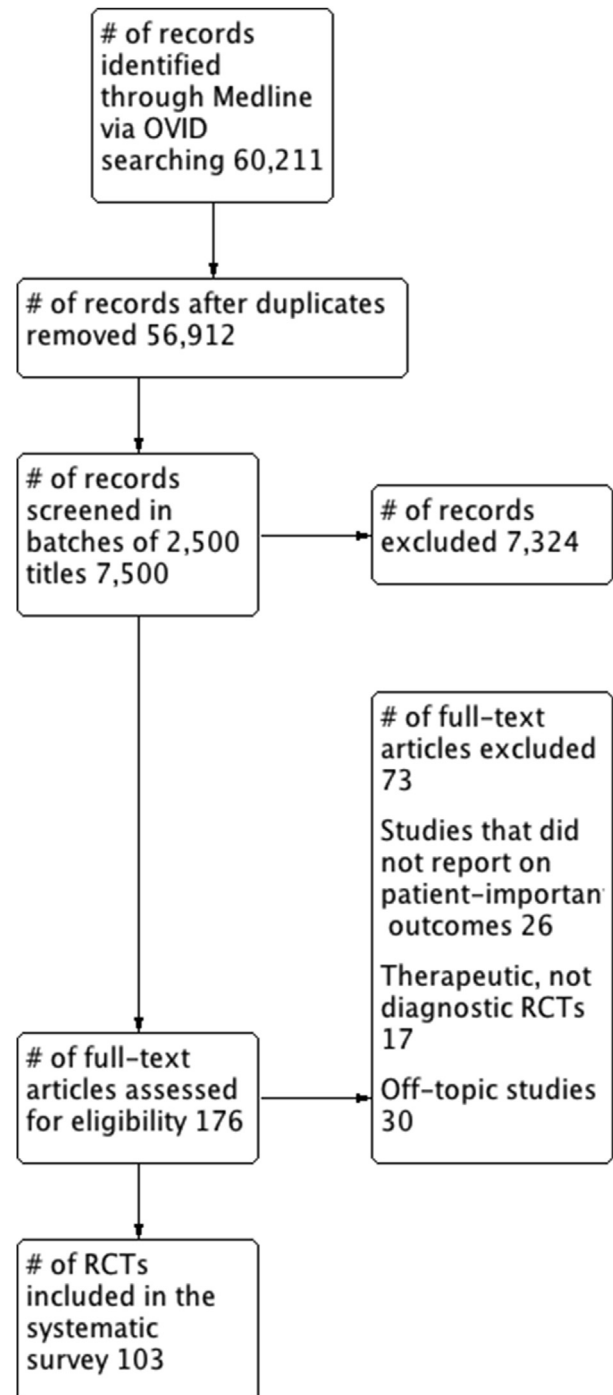


Fig. 1. Study flow diagram. RCTs, randomized clinical trials.

Studies were most often conducted in Europe ($n = 44$; 42.7%) or in North America ($n = 34$; 33.0%) and seldom in Australia and New Zealand ($n = 3$; 2.9%). Twenty-two studies (21.3%) did not report where the study was conducted. The median sample size was 248 patients (interquartile range: 111.5–723); 14.6% had sample sizes of more than 500 patients (Fig. 4).

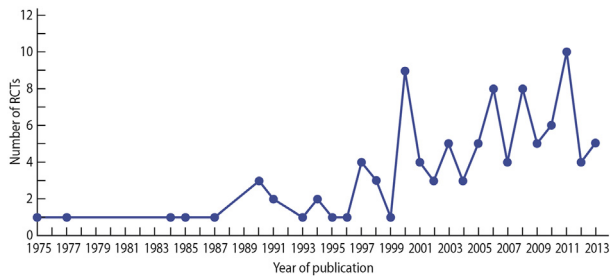


Fig. 2. Distribution of year of publication. RCTs, randomized clinical trials.

Most studies addressed issues of diagnostic imaging ($n = 72$; 69.9%) or laboratory tests ($n = 22$; 21.4%) (Fig. 5). In 19 RCTs (18.4%), the control test was an aspect of physical examination; 24 (23.3%) of the control group arms did not include any comparator test (Appendix Fig. 2 on the journal's Web site at www.elsevier.com). Of the 103 eligible studies, 17 (16.5%) implemented a strict protocol linking diagnostic test results to a specific clinical management strategy for both intervention and control groups. Two other studies (1.9%) implemented protocolized management for the intervention but not the control arm. The majority ($n = 60$, 58.2%) reported sample size calculations; however, only nine (15.0%) were powered to detect relative risk reductions of 35% or less.

3.3. Risk of bias

We found limitations of risk of bias very similar in studies that reported statistically significant results and those that did not. In both sets of studies, we judged approximately half as low risk of bias with regard to missing outcome data and random sequence generation and a third or less at low risk of bias regarding allocation concealment and blinding (Table 1).

3.4. Outcomes

Of the 103 RCTs, 41 (39.8%) reported on mortality, 63 (61.2%) on morbidities, 14 (13.6%) on symptoms/QoL/

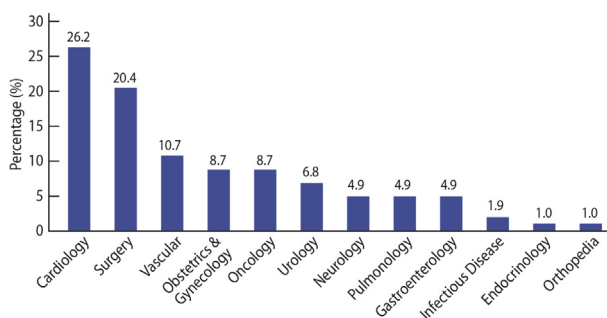


Fig. 3. Percentage of included study per medical content area.

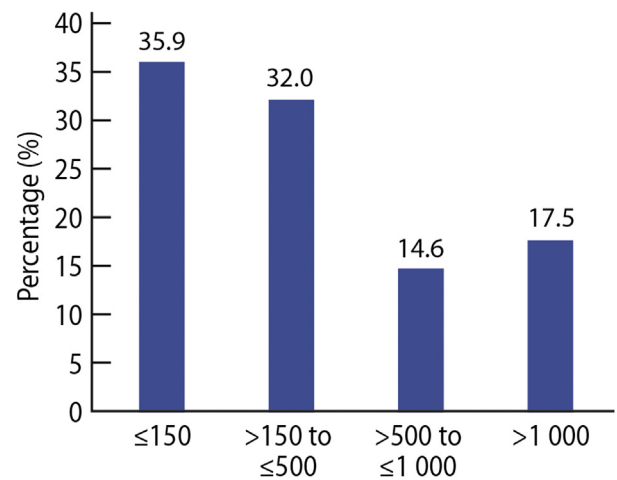


Fig. 4. Distribution of sample sizes.

functional status, 10 (9.7%) on composite end points, 41 (39.8%) on surrogate outcomes, and 24 (23.3%) on costs.

3.5. Trials' results

3.5.1. Mortality

Of the 41 (39.8%) RCTs that evaluated mortality, only one (2.4%) [20], a trial of positron emission tomography with 18-fluorodeoxyglucose compared to conventional staging for diagnosis of suspected non-small cell lung cancer, reported statistically significant results (RR = 0.33, 95% CI: 0.14–0.79). Of the 40 (97.6%) RCTs with a nonstatistically significant results, only two (5.0%) [21,22] were adequately powered.

3.5.2. Morbidity

Of the 63 (61.2%) RCTs that addressed morbidity, 11 (17.5%) [20,23–32] reported statistically significant results (Table 2). All [20,23–32] reported lower morbidity in the experimental group (relative risk lower than 0.8 in 10,

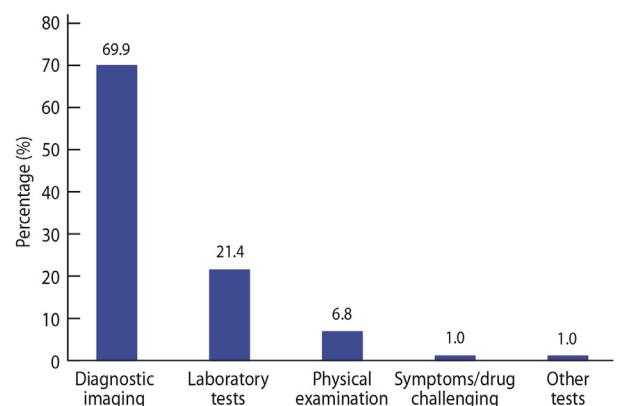


Fig. 5. Major category for sorts of tests in the intervention arm.

Table 1. Risk of bias assessment comparing studies that reported statistically significant results with those that did not report a statistically significant result

<i>N (%)</i>	Was the allocation sequence adequately generated?	Was allocation adequately concealed?	Blinding: was knowledge of the allocated interventions adequately prevented?	Was loss to follow-up (missing outcome data) infrequent?	Are reports of the study free of suggestion of selective outcome reporting?	Was the study apparently free of other problems (i.e., conflict of interest) that could put it at a risk of bias?
Definitely yes (low risk of bias)	6 (50.0) <i>46 (50.0)</i>	1 (8.3) <i>34 (36.9)</i>	4 (33.3) <i>18 (19.6)</i>	7 (58.3) <i>51 (55.4)</i>	4 (33.3) <i>33 (36.3)</i>	6 (50.0) <i>28 (30.8)</i>
Probably yes	1 (8.3) <i>12 (13.0)</i>	3 (25.0) <i>15 (16.3)</i>	2 (16.7) <i>15 (16.3)</i>	4 (33.3) <i>20 (21.7)</i>	7 (58.3) <i>43 (46.7)</i>	5 (41.7) <i>42 (45.6)</i>
Probably no	4 (33.3) <i>27 (29.7)</i>	4 (33.3) <i>29 (31.5)</i>	1 (8.3) <i>28 (30.4)</i>	0 (0.0) <i>10 (10.9)</i>	0 (0.0) <i>10 (10.9)</i>	0 (0.0) <i>17 (18.5)</i>
Definitely no (high risk of bias)	1 (8.3) <i>6 (6.5)</i>	4 (33.3) <i>13 (14.3)</i>	5 (41.7) <i>30 (33.0)</i>	1 (8.3) <i>10 (11.0)</i>	1 (8.3) <i>5 (5.4)</i>	1 (8.3) <i>4 (4.3)</i>

Studies that reported statistically significant results ($n = 12$) are shown in the first line.

Studies that did not report a statistically significant result ($n = 91$) are shown in italics in the second line.

relative risk greater than 1.2 in 1 [32]) (Table 2). Of the 52 (82.5%) RCTs with a nonstatistically significant results, 10 (19.2%) RCTs were adequately powered.

3.5.3. Symptoms/QoL/functional status

Of 14 (13.6%) RCTs that evaluated symptoms/QoL/functional status all reported results as a binary outcome of which only one (7.1%) [23], a trial of esophageal Doppler-guided cardiovascular optimization compared to standard intraoperative fluids for patients undergoing radical cystectomy evaluating nausea and vomiting, reported statistically significant results (RR = 0.08, 95% CI: 0.01–0.59). Of the 13 (92.8%) RCTs with a nonstatistically significant results, 3 (23.1%) RCTs were adequately powered.

3.5.4. Composite end points

Of 10 (9.7%) RCTs that designated a composite outcome, one (10.0%) [33], a trial of continuous 2-MHz transcranial Doppler ultrasonography compared to placebo for acute ischemic stroke evaluating complete recanalization or dramatic clinical recovery within 2 hours after the administration of a t-PA bolus, reported statistically significant results (RR = 1.63, 95% CI: 1.04–2.56) favoring the intervention group. Of the nine (90.0%) RCTs with a nonstatistically significant results, no RCT was adequately powered.

4. Discussion

4.1. Main findings

Our random sample of approximately 13% of the records identified through the search of the MEDLINE database identified 103 RCTs of diagnostic tests focusing on patient-important outcomes (Fig. 1), suggesting that MEDLINE includes approximately 781 such RCTs. Other databases will include additional RCTs. Of the RCTs we included, almost all were published in

1997 or later, although from 2000 onward, the number published each year has been very similar (Fig. 2). The most common areas of investigation were cardiology (26.2%) and surgery (20.4%) (Fig. 3) and the types of tests diagnostic imaging (69.9%) and laboratory tests (21.3%) in the intervention arms (Fig. 5).

Of the 103 trials, 12 (11.6%) reported statistically significant differences between groups. Positive studies examined a wide variety of alternative diagnostic strategies. These included additional imaging in comparison to conventional diagnostic approaches [20,23,24,26–28,32]; alternative or additional physiologic measurement [29]; alternative techniques for tissue specimens [25]; and alternative provocative tests [30,31] (Table 2). These studies highlight the feasibility of designing and carrying out RCTs that establish the superiority of one diagnostic strategy vs. another on patient-important outcomes.

Of the studies that failed to report statistically significant difference, relatively few (4.9% reporting on mortality, 16.9% on morbidity, and 23.5% on function) were powered to detect modest but important treatment effects. Many studies were also limited with respect to risk of bias: approximately 50% failed to adequately conceal randomization, and 60% failed to address blinding. These limitations existed to a similar extent in studies that did, and did not, report statistically significant differences between groups (Table 1). Although blinding of patients and clinicians may often be impossible in diagnostic test RCTs, it should almost always be possible to blind outcome assessors and data analysts.

4.2. Strengths and limitations

Ours is the first systematic exploration of the frequency and characteristics of RCTs of alternative diagnostic test strategies. Strengths of our review include a comprehensive

Table 2. Positive studies for morbidities

Study identifier	PIC	Morbid outcome	Intervention and control groups (number of events per total number of participants)	Risk relative (95% CI)
van Tinteren et al. 2002 [20]	P = non–small cell lung cancer I = positron emission tomography (PET) with 18-fluorodeoxyglucose (18FDG) C = conventional staging	Futile thoracotomy	Conventional workup and PET (CWU + PET): 19/92 Conventional workup: 39/96	0.51 (0.32–0.81)
Pillai et al. 2011 [23]	P = undergoing radical cystectomy I = esophageal Doppler-guided cardiovascular optimization C = standard intraoperative fluids	Ileus	Doppler: 7/32 Standard: 18/34	0.41 (0.20–0.86)
McKenna et al. 2003 [24]	P = low risk at 30-week gestation I = ultrasound scan + standard antenatal care C = standard antenatal care	Infants assessed as small for age at birth	Ultrasound: 69/994 Standard care: 104/999	0.67 (0.50–0.89)
Tournoy et al. 2008 [25]	P = proven or suspected non–small cell lung cancer I = real time–guided fine-needle aspiration (EUS-FNA) C = appropriate surgical staging procedure	Thoracotomy for surgical staging	EUS-FNA: 6/19 Surgical staging procedure: 21/21	0.33 (0.18–0.63)
Fischer et al. 2009 [26]	P = non–small cell lung cancer I = conventional staging and combined positron emission tomography and computed tomography (PET–CT) C = conventional staging alone	Number of futile thoracotomies in patients who had recurrent disease	PET–CT: 3/60 Conventional: 13/73	0.28 (0.08–0.94)
van Loon et al. 1997 [27]	P = women with breech presentation at term I = magnetic resonance (MR) pelvimetry (results reported to obstetricians) C = MR pelvimetry (results were not disclosed, and decisions were made on the basis of clinical factors)	Emergency caesarean-section	Results reported: 22/118 Results not reported: 41/117	0.53 (0.34–0.83)
Klein et al. 2001 [28]	P = atrial fibrillation who are to undergo electrical cardioversion I = transesophageal echocardiography (TE) C = conventional treatment	Hemorrhagic events	TE: 18/619 Conventional: 33/603	0.53 (0.30–0.93)
Hosking et al. 1987 [29]	P = esophageal varices I = intravariceal sclerotherapy (IS) based on manometric assessment C = IS based on visual assessment	Upper gastrointestinal bleeds from esophageal varices	Manometric: 1/33 Visual: 14/35	0.08 (0.01–0.54)
Höjer et al. 1990 [30]	P = coma of unclear origin with suspected poisoning I = benzodiazepine antagonist flumazenil C = placebo	Gastric lavage	Flumazenil: 21/53 Placebo: 31/52	0.66 (0.45–0.99)
Okajima et al. 2006 [31]	P = coronary artery disease I = myocardial contrast echocardiography (MCE) with nicorandil infusion C = MCE with dipyridamole	Minor adverse reactions (anginal symptoms or arrhythmia, flushing, sensations of warmth, headaches)	Nicorandil: 0/44 Dipyridamole: 29/44	0.02 (0.00–0.27)
Goodacre et al. 2011 [32]	P = acute chest pain due to suspected myocardial infarction I = panel assay of cardiac markers C = standard care (SoC)	Successful discharge from hospital	Panel: 358/1,125 SoC: 146/1,118	2.44 (2.05–2.90)

Abbreviations: CI, confidence interval; NR, not reported; PIC, patient; intervention; and control group.

search; independent and duplicate assessment of eligibility, risk of bias, and data abstraction; and a focus on outcomes important to patients. Our goal was not to be comprehensive or necessarily fully representative but rather to provide initial insights into the nature and extent of RCTs of diagnostic tests focusing on patient-important outcomes. Our search was therefore comprehensive within the MEDLINE database, though restriction to MEDLINE represents a limitation. Diagnostic test RCTs published in other databases may differ in areas of focus and results.

Diagnostic tests can only influence outcome if the results are clearly linked to beneficial management strategies. One might therefore consider the studies' uniform failure to establish a protocolized response to diagnostic test results as a limitation. On the other hand, a study that established such a protocol would move toward the explanatory or mechanistic end of the mechanistic-practical trial spectrum [34]. The performance of practical trials could also be seen as strength of the existing studies.

4.3. Implications

The explosion of sophisticated medical technologies has brought enormous benefits with substantial impact on reduction in population morbidity and improvements in QoL. The proliferation, particularly in the area of diagnostic tests, has also raised increasing concern regarding overdiagnosis and overtreatment [35–37]. There are circumstances in which a test results is unequivocally linked to an intervention in which high-quality evidence has established either benefit or harm reduction. Under such circumstances, RCTs of diagnostic tests are not necessary [10]. There are, however, many other circumstances (likely the majority of testing situations) in which such links do not exist. In this latter set of circumstances, insisting on compelling demonstration of patient-important benefit from expensive or invasive testing procedures represents one strategy for addressing this concern.

Most studies of diagnostic technologies have failed to demonstrate differences in patient-important outcomes. One possible conclusion, highly relevant to the overdiagnosis issue, is that little of importance to patients is gained from most tests that carry with them additional expense and burden.

There are, however, alternative explanations for the large number of negative results. One might be that RCTs of diagnostic tests present excessive challenges, and we cannot expect positive findings. Our results are not consistent with that explanation: 12 of 103 RCTs demonstrated benefits of one diagnostic strategy over a comparator, a relatively small but still important yield. Had a larger proportion of RCTs been adequately powered, one would anticipate additional positive results.

Other alternatives are of more concern. Most negative studies were underpowered for modest but potentially important effects and could represent false negatives. High risk of bias could also be a concern, though it is likely that

high risk of bias more often results in spurious effects than in failure to detect effects that are in fact present.

Overall, our results demonstrate the feasibility of RCTs of diagnostic tests, and their potential, when adequately powered and rigorously planned and carried out, of demonstrating benefit of one strategy over another on patient-important outcomes. This potential will only be realized, however, when the importance of diagnostic test RCTs is fully recognized, and the requisite resources and expertise are applied to ensure their optimal planning and implementation.

Acknowledgments

The authors would like to thank Neera Bhatnagar for her help with the search strategy.

Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.jclinepi.2016.12.009>.

References

- [1] Simel DL, Rennie D. *The Rational Clinical Examination: Evidence-Based Clinical Diagnosis*. New York: McGraw-Hill; 2009.
- [2] Pines J, Carpenter CR, Raja A, Schuur J. *Evidence-Based Emergency Care: Diagnostic Testing and Clinical Decision Rules*. 2nd ed. Oxford UK: Wiley-Blackwell; 2013.
- [3] Bossuyt PM, Reitsma JB, Linnet K, Moons KG. Beyond diagnostic accuracy: the clinical utility of diagnostic tests. *Clin Chem* 2012; 58:1636–43.
- [4] Peeling RW, Smith PG, Bossuyt PM. A guide for diagnostic evaluations. *Nat Rev Microbiol* 2010;8(12 Suppl):S2–6.
- [5] Moons KG, de Groot JA, Linnet K, Reitsma JB, Bossuyt PM. Quantifying the added value of a diagnostic test or marker. *Clin Chem* 2012;58:1408–17.
- [6] Guyatt GH, Tugwell PX, Feeny DH, Haynes RB, Drummond M. A framework for clinical evaluation of diagnostic technologies. *CMAJ* 1986;134(6):587–94.
- [7] Guyatt GH, Tugwell PX, Feeny DH, Drummond MF, Haynes RB. The role of before-after studies of therapeutic impact in the evaluation of diagnostic technologies. *J Chronic Dis* 1986;39(4):295–304.
- [8] Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;11:88–94.
- [9] Kanzaria HK, McCabe AM, Meisel ZM, LeBlanc A, Schaffer JT, Bellolio MF, et al. Advancing patient-centered outcomes in emergency diagnostic imaging: a research Agenda. *Acad Emerg Med* 2015;22(12):1435–46.
- [10] Lord SJ, Irwig L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med* 2006;144:850–5.
- [11] Guyatt GH, Montori VM, Devereaux PJ, Schünemann HJ, Bhandari M. Patients at the center: in our practice, and in our use of language (Editorial). *ACP J Club* 2004;140:A11.
- [12] Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al, GRADE Working Group. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008;336:1106–10.
- [13] Bloom SL, Spong CY, Thom E, Varner MW, Rouse DJ, Weininger S, et al. Fetal pulse oximetry and cesarean delivery. *N Engl J Med* 2006; 355:2195–202.

- [14] Moe GW, Howlett J, Januzzi JL, Zowall H, Canadian Multicenter Improved Management of Patients With Congestive Heart Failure (IMPROVE-CHF) Study Investigators. N-terminal pro-B-type natriuretic peptide testing improves the management of patients with suspected acute heart failure: primary results of the Canadian prospective randomized multicenter IMPROVE-CHF study. *Circulation* 2007;115:3103–10.
- [15] Laine L, Sahota A, Shah A. Does capsule endoscopy improve outcomes in obscure gastrointestinal bleeding? Randomized trial versus dedicated small bowel radiography. *Gastroenterology* 2010;138(5):1673–1680.e1. quiz e11–2.
- [16] Miller CD, Hwang W, Hoekstra JW, Case D, Lefebvre C, Blumstein H, et al. Stress cardiac magnetic resonance imaging with observation unit care reduces cost for patients with emergent chest pain: a randomized trial. *Ann Emerg Med* 2010;56(3):209–219.e2.
- [17] Calhoun WJ, Ameredes BT, King TS, Icitovic N, Bleecker ER, Castro M, et al. Comparison of physician-, biomarker-, and symptom-based strategies for adjustment of inhaled corticosteroid therapy in adults with asthma: the BASALT randomized controlled trial. *JAMA* 2012;308:987–97.
- [18] Akl EA, Briel M, You JJ, Lamontagne F, Gangji A, Cukierman-Yaffe T, et al. LOST to follow-up Information in Trials (LOST-IT): a protocol on the potential impact. *Trials* 2009;10:40.
- [19] Guyatt GH, Busse JW. Modification of Cochrane Tool to assess risk of bias in randomized trials. Available at <http://distillercer.com/resources/>. Accessed February 4, 2015.
- [20] van Tinteren H, Hoekstra OS, Smit EF, van den Bergh JH, Schreurs AJ, Stallaert RA, et al. Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS multicentre randomised trial. *Lancet* 2002;359:1388–93.
- [21] Rozanski A, Gransar H, Shaw LJ, Kim J, Miranda-Peats L, Wong ND, et al. Impact of coronary artery calcium scanning on coronary risk factors and downstream testing the EISNER (Early Identification of Sub-clinical Atherosclerosis by Noninvasive Imaging Research) prospective randomized trial. *J Am Coll Cardiol* 2011;57:1622–32.
- [22] Davies AH, Hawdon AJ, Sydes MR, Thompson SG, VGST Participants. Is duplex surveillance of value after leg vein bypass grafting? Principal results of the Vein Graft Surveillance Randomised Trial (VGST). *Circulation* 2005;112:1985–91.
- [23] Pillai P, McEleavy I, Gaughan M, Snowden C, Nesbitt I, Durkan G, et al. A double-blind randomized controlled clinical trial to assess the effect of Doppler optimized intraoperative fluid management on outcome following radical cystectomy. *J Urol* 2011;186:2201–6.
- [24] McKenna D, Tharmaratnam S, Mahsud S, Bailie C, Harper A, Dornan J. A randomized trial using ultrasound to identify the high-risk fetus in a low-risk population. *Obstet Gynecol* 2003;101(4):626–32.
- [25] Tournoy KG, De Ryck F, Vanwalleghem LR, Vermassen F, Praet M, Aerts JG, et al. Endoscopic ultrasound reduces surgical mediastinal staging in lung cancer: a randomized trial. *Am J Respir Crit Care Med* 2008;177:531–5.
- [26] Fischer B, Lassen U, Mortensen J, Larsen S, Loft A, Bertelsen A, et al. Preoperative staging of lung cancer with combined PET-CT. *N Engl J Med* 2009;361:32–9.
- [27] van Loon AJ, Mantingh A, Serlier EK, Kroon G, Mooyaart EL, Huisjes HJ. Randomised controlled trial of magnetic-resonance pelvimetry in breech presentation at term. *Lancet* 1997;350:1799–804.
- [28] Klein AL, Grimm RA, Murray RD, Apperson-Hansen C, Asinger RW, Black IW, et al. Assessment of Cardioversion Using Transesophageal Echocardiography Investigators. Use of transesophageal echocardiography to guide cardioversion in patients with atrial fibrillation. *N Engl J Med* 2001;344:1411–20.
- [29] Hosking SW, Robinson P, Johnson AG. Usefulness of manometric assessment of varices in maintenance sclerotherapy. A controlled trial. *Gastroenterology* 1987;93(4):846–51.
- [30] Höjer J, Baehrendtz S, Matell G, Gustafsson LL. Diagnostic utility of flumazenil in coma with suspected poisoning: a double blind, randomised controlled study. *BMJ* 1990;301:1308–11.
- [31] Okajima K, Kawase Y, Matsushita N, Iwata S, Doi A, Hasegawa T, et al. Usefulness of myocardial contrast echocardiography with nicorandil stress for the detection of coronary artery stenosis. *Heart* 2006;92:1331–2.
- [32] Goodacre SW, Bradburn M, Cross E, Collinson P, Gray A, Hall AS, et al. The Randomised Assessment of Treatment using Panel Assay of Cardiac Markers (RATPAC) trial: a randomised controlled trial of point-of-care cardiac markers in the emergency department. *Heart* 2011;97:190–6.
- [33] Alexandrov AV, Molina CA, Grotta JC, Garami Z, Ford SR, Alvarez-Sabin J, et al. Ultrasound-enhanced systemic thrombolysis for acute ischemic stroke. *N Engl J Med* 2004;351:2170–8.
- [34] Karanickolas PJ, Montori VM, Devereaux PJ, Schünemann H, Guyatt GH. A new “mechanistic-practical” framework for designing and interpreting randomized trials. *J Clin Epidemiol* 2009;62:479–84.
- [35] Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. *Lancet* 2014;383:156–65.
- [36] Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, et al. Biomedical research: increasing value, reducing waste. *Lancet* 2014;383:101–4.
- [37] Glasziou P, Moynihan R, Richards T, Godlee F. Too much medicine; too little care. *BMJ* 2013;347:f4247.